

DYNAMIC BILEVEL THRESHOLDING OF DIGITAL IMAGES

FIELD OF INVENTION

5 This invention relates generally to automated image analysis and more specifically to segmentation of images by thresholding.

BACKGROUND OF THE INVENTION

10 Image segmentation is a process of dividing an image into regions of interest, and includes discrimination of objects in an image from a background. Thresholding is one technique for segmentation. In bilevel thresholding, each image pixel is assigned to one of two classes according to whether its intensity (gray level or color) is greater or less than a specified threshold, resulting in a binary image. In multilevel thresholding, the entire image is thresholded multiple times, each time with a different constant threshold, resulting in multiple binary images.

15 One common application is optical character recognition (OCR), where image pixels are typically segmented into characters by thresholding. Consider, for example, an image of black text on a white background. A histogram of all the intensity values in the image will have two dominant peaks: one peak corresponding to the intensity value of the black text, and a second peak corresponding to the intensity value of the white background. If a threshold is set at an intensity value that is in the bottom of the valley between the two peaks, then any pixel having an intensity value darker than the threshold may be assigned to text, and any pixel
20 having a intensity value lighter than the threshold may be assigned to background.

25 In the case of black text on a white background, a constant global threshold may be determined from a intensity value histogram of the entire image. However, many images of interest are more complex than just black text against a white background. For example, an image may include blocks of color (that is, the background may vary), text may overlap blocks of different colors, and text may be
30

lighter than the local background. For complex images, the threshold may be dynamic, varying depending on the location of the pixel of interest within the image. A dynamic threshold may be dependent on intensity value data over a region of an image, or a dynamic threshold may vary from pixel to pixel. See, for example, Joan S. Weszka, "A Survey of Threshold Selection Techniques", *Computer Vision, Graphics, and Image Processing* 7, 259-265 (1978) and Sahoo *et al.*, "A Survey of Thresholding Techniques", *Computer Vision, Graphics, and Image Processing* 41, 233-260 (1988).

Particular problems for thresholding include determination of a suitable threshold at the boundaries of objects, determination of a suitable threshold for thin objects (where there are few object intensity values in the histogram), and determination of a suitable threshold when there are areas of interest that are lighter than the background.

There is a need for improved segmentation of complex images using bilevel thresholding.

SUMMARY OF THE INVENTION

For each pixel, a threshold is selected from a set of thresholds. In a first example embodiment, at least one threshold is variable, and one threshold is a constant value.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a gray-scale image with regions of interest to illustrate bilevel thresholding resulting from several example thresholding methods.

Figure 2 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1, using a dynamic threshold that is constant within blocks (or cells, or tiles).

5 Figure 3 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1, using a dynamic threshold that changes from pixel to pixel.

10 Figure 4 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1, using a set of thresholds in accordance with an example embodiment of the invention.

15 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT OF THE INVENTION

20 Figure 1 illustrates an image having a uniform gray background (depicted by cross-hatching) with a solid black area 100, a solid white area 102, and a relatively thin black line 104 near the solid white area 102. Note that the white area 102 does not have a black line border around it; it is simply a white area within the surrounding gray background. In the following discussion, two prior-art methods of bilevel thresholding are applied to an image like figure 1 with a uniform gray background, and the resulting binary images are illustrated (figures 2 and 3). Then an example embodiment of a method in accordance with the invention is applied to an image like figure 1 with a uniform gray background, and the resulting binary image is illustrated (figure 4). It is important to note that due to limitations on what is permitted for patent illustrations, the gray background of figure 1 is not uniform, and is simulated by cross-hatching. Figures 2-4 illustrate the result of applying various bilevel thresholding algorithms to an image like figure 1 but having a

uniform gray background, and do not illustrate the result of applying the same thresholding algorithms literally to figure 1 with its simulated gray background.

Note also in the following discussion that it is assumed that low intensity pixels have low numerical intensity values and that high intensity pixels have high numerical intensity values. This may be reversed, so that low intensities are represented by high numbers and vice versa, in which case the MAX functions become MIN functions, signs are reversed, and so forth.

Figure 2 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1 (with a uniform background), using a thresholding method that is typical for some commercially available OCR software. The following is a simplified description of the general type of thresholding involved in producing figure 2, and may not correspond precisely to any particular commercially available software. An additional example may be found in U.S. Patent Number 5,651,077,

For the method illustrated in figure 2, the overall image of figure 1 is partitioned into blocks (also called cells, or tiles), for example, square blocks of 64x64 pixels. A co-occurrence matrix is computed for each block. Assume for simplicity of illustration that gray-levels only have 4 bits (gray levels 0-15). The co-occurrence matrix, may be, for example, a 16x16 array. An entry in the array for position (i,j) is the frequency of occurrence of adjacent pixels with gray levels i and j. A histogram is generated from the co-occurrence matrix including only those entries close to the diagonal. For blocks with distinct bi-modal histograms (for example, dark text on a light background) the threshold is computed to be near the histogram valley. For blocks without a distinct bi-modal histogram, for example, a unimodal histogram for a block within a background area, may be assigned a threshold similar to a nearby block with similar background and previously computed threshold.

When a threshold is computed for a block, the threshold may change significantly from one block to the next, sometimes resulting in block-sized artifacts. Blocks that include parts of white area 102 in figure 1 have a higher intensity threshold relative to blocks comprising only the background gray-level. As a result, the threshold near white area 102 may be greater than the intensity of the gray background, so that the gray-levels in areas near white area 102 in figure 1 may snap to black in the binary image, as illustrated by the wide black areas around area 202 in figure 2. As a result, a wide black frame is generated as an artifact in figure 2, and the black line 104 in figure 1 is completely lost in the binary image of figure 2.

Figure 3 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1 (with a uniform background), using a thresholding method that varies the threshold on a pixel by pixel basis. For the particular example of figure 3, for each pixel, the threshold is the maximum intensity value among the KxK surrounding pixels, less a constant offset. This technique sets the threshold below the background for each pixel. In addition, the technique behaves as a high-pass filter, or edge detector. This has both advantages and disadvantages. Advantages include: (1) light areas are preserved as relatively narrow black frames with white interiors, without the artifacts of block-based methods, and (2) thin objects such as lines are preserved. A disadvantage is that large dark areas are reduced to frames with white interiors, rather than being preserved as dark areas.

Consider, for example, pixels near area 102 in figure 1. First consider a pixel where the KxK surrounding area includes only gray background pixels. The threshold is set to an offset below the intensity of the background, and the pixel under consideration is snapped to white. Now consider a pixel that has a background level intensity, but the KxK surrounding area includes part of area 102. Now the threshold is set to an offset just below white, and the pixel under

consideration is snapped to black. Finally, consider a pixel where the KxK surrounding area includes only the white pixels of area 102. The threshold is again an offset just below white, and the pixel under consideration is snapped to white. As a result, in figure 3, the background snaps to white, and a black line results at each transition from gray to white and vice versa, thereby distinguishing area 302 within the white background in figure 3. Note also that the line 304 in figure 3, corresponding to line 104 in figure 1, is also distinguished.

Now consider pixels near area 100 in figure 1. Again, the background snaps to white. Consider a pixel just inside area 100, where part of the gray background is included in the KxK surrounding area. The threshold is set to an offset just below the gray background, and the black pixel under consideration is snapped to black. Now consider a pixel inside area 100 where the KxK surrounding area includes only black pixels. The threshold is set to an offset below black (or to a low intensity limit), and the black pixel under consideration is snapped to white. As a result, a solid black area 100 in figure 1 is rendered as a black frame with a white interior in figure 3. For some segmentation requirements, it may be preferable to render large dark areas in the gray-level image as black areas in the binary segmented image.

Figure 4 is a binary image illustrating the result of applying a bilevel thresholding method to the image of figure 1 (with a uniform background), using a thresholding method in accordance with an example embodiment of the invention. In the example embodiment illustrated by figure 4, the threshold is determined on a pixel by pixel basis. The threshold is selected from multiple thresholds, at least one of which is dynamic. The thresholding method has particular advantages in OCR, but is not limited to OCR. In the example embodiment illustrated by figure 4, the threshold T for a pixel at row w , column c , is determined according to the following equation:

$$T(r,c) = \text{MAX}[\text{MAX}_K(r,c)-T1, T2] \quad \text{Equation 1}$$

where:

$\text{MAX}_K(r,c)$ is the maximum intensity value of the $K \times K$ pixels surrounding pixel (r,c) .

$T1$ is an intensity offset value, which may be a constant value.

$T2$ is an intensity value, which may be a constant value for an entire image.

Note that in the example of equation 1, the threshold T is selected from the highest intensity of two thresholds, one of which is dynamic and one of which is constant.

Consider the application of equation 1 to figure 1 (with a uniform background). Assume that $T2$ is an intensity that is lower than the background intensity minus $T1$. For all of figure 1 other than inside area 100, the variable threshold ($\text{MAX}_K(r,c)-T1$) will always be higher than $T2$. Note that other than inside area 100, the variable threshold is either the background intensity less $T1$, or is the white intensity less $T1$. Inside area 100, for pixels where the $K \times K$ surrounding pixels are all black, the variable threshold is less than $T2$, and T is selected to be $T2$. Accordingly, black pixels within area 100 snap to black, and the resulting binary image is as illustrated in figure 4.

When using Equation 1, in the binary segmented image, the frames around light areas will be approximately K pixels wide. K may be as small as $K=2$. If the input sampling rate is 300 pixels per inch, a suitable example value for K for OCR is $K=7$. $T1$ or $T2$ may vary by region or block, but remain constant within a region or block. However, there is a risk of artifacts at block boundaries. $T1$ needs to be large enough to ensure that $[\text{MAX}_K(r,c)-T1]$ is well below the background to minimize noise. Accordingly, a suitable example value for $T1$ is about 30% of the intensity range for the overall image. A more accurate approach is to make $T1$ a function of the spread of the background. For example, in an histogram of the

overall image, there may be a peak in the dark area caused by the background. T1 may then be made a multiple of the standard deviation of the data in the peak in the dark area, for example, twice the standard deviation. T2 may be determined based on the overall image, using any of the known techniques for determining a single threshold.

The foregoing description of the present invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and other modifications and variations may be possible in light of the above teachings. The embodiment was chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and various modifications as are suited to the particular use contemplated. It is intended that the appended claims be construed to include other alternative embodiments of the invention except insofar as limited by the prior art.